

DOCUMENT RESUME

ED 306 294

TM 013 164

AUTHOR Schuldberg, David
TITLE Indices of Individuals' Sensitivities To Computerized Test Administration and Repeated Testing.
PUB DATE 5 Oct 88
NOTE 32p.; Sponsored by grants from Montanans on a New Track for Science (MONTs) and the University of Montana Office of Research Administration.
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS College Students; *Computer Assisted Testing; Higher Education; *Individual Differences; *Objective Tests; Responses; *Test Format; Test Reliability
IDENTIFIERS *Minnesota Multiphasic Personality Inventory; Paper and Pencil Tests; Repeated Testing; *Testing Effects; Test Retest Reliability

ABSTRACT

Indices were constructed to measure individual differences in the effects of the automated testing format and repeated testing on Minnesota Multiphasic Personality Inventory (MMPI) responses. Two types of instability measures were studied within a data set from the responses of 150 undergraduate students who took a computer-administered and pencil-and-paper MMPI a week apart. Two subject groups included 42 males and 33 females each. One set of indices measured systematic format- and time-related changes in responding, shifting attributable to format or time alone. Two families of six indices each were computed measuring unsystematic changes in responding, or overall tendencies to shift in a particular direction among the responses "true," "false," and "cannot say." These unsystematic changes were assessed both between formats and across times, although they were partially confounded in the present study. Systematic format shifting was related to a more general and unsystematic tendency to shift between "true" and "false" responses. The use of "cannot say" in the computerized testing situation appears distinct from the tendency to use the "cannot say" response on the pencil-and-paper test. Systematic item shifting attributable to time, although not involving an internally consistent set of responses, is distinct from other instability indices derived in this study and is therefore sensitive to the design of the administration software. Personality and other correlates of the item-shifting indices are discussed. Five tables present study data. (Author/SLD)

* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *

ED306294

(October 5, 1988)

Indices of individuals' sensitivities to computerized
test administration and repeated testing

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

- ☒ This document has been reproduced as
received from the person or organization
originating it.
☐ Minor changes have been made to improve
reproduction quality.

- Points of view or opinions stated in this docu-
ment do not necessarily represent official
OERI position or policy.

David Schuldberg

Department of Psychology

University of Montana

and

Department of Psychiatry

Yale University

"PERMISSION TO REPRODUCE THIS
MATERIAL HAS BEEN GRANTED BY

DAVID SCHULDBERG

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)."

A shorter version of this paper was presented at the 96th. Annual
Convention of the American Psychological Association, Atlanta,
Ga., August 12-16, 1988.

RUNNING HEAD: Sensitivities to format and repeated testing

TMO 13164

Abstract

Indices are constructed to measure individual differences in the effects of the automated testing format and repeated testing on MMPI responses. Two types of instability measure are studied within a data set from the responses of 150 undergraduate subjects who took a computer-administered and Pencil and paper MMPI a week apart. One set of indices measures systematic format- and time-related changes in responding, shifting attributable to format or time alone. Two families of six indices each are computed measuring unsystematic changes in responding, overall tendencies to shift in a particular direction between the particular responses, "True", False", and "Cannot Say". These Unsystematic changes are assessed both between formats and across times, although these factors are partially confounded in the present study. Systematic Format shifting is related to a more general, unsystematic tendency to shift between "True" and "False" responses. The use of "Cannot Say" in the computerized testing situation appears distinct from the tendency to use "Cannot Say" on the Pencil and Paper test. Systematic item shifting attributable to Time, although not involving an internally consistent set of responses, is distinct from the other instability indices derived in this study. Personality and other correlates of the item-shifting indices are discussed.

Indices of individuals' sensitivities to computerized
test administration and repeated testing

This paper investigates individual differences in the effects of automated test administration and repeated testing on subjects' responses to the Minnesota Multiphasic Personality Inventory (MMPI; Hathaway and McKinley, 1967). The effects of different assessment formats (e.g., Computer-administered vs. Pencil and paper), as well as the effects of taking an instrument more than once, may be the same for everyone who takes a test. Or, particular people may be sensitive to different administrations in particular ways. Differences have been found in test results and client attitudes related to computerized tests (Ben-Porath and Butcher, 1987), and it is important to evaluate format equivalence and the size of mode or format differences when using a computerized test (Butcher, 1987; Honaker, 1987) with a particular patient.

Research to date has indicated that the effects of computerized testing are relatively small. The early prediction of increased candor and less defensiveness in computerized assessment has not been borne out in research on objective personality tests. A number of studies using MMPI scales have employed a variety of designs to test this hypothesis and to examine format equivalence in general (Bresolin, 1984; Biskin and Kolotkin, 1977; Evan and Miller, 1969; Hart and Goldstein, 1985; Koson, Kitchen, Kochen and Stodolosky, 1970; Lambert et al 1987; Lushene, O'Neill, and Dunn, 1974; Russell, Peace, and Mellsop, 1986; Schuldborg, in press; White, Clements and Fowler, 1985). Overall, computerized administration tends to produce less

elevated MMPI profiles than traditional procedures, even when test takers' use of the "Cannot Say" or unscorable response is controlled. These format effects have tended to be small, although differences emerge when robust designs are employed. Research has almost exclusively dealt with scale rather than item equivalence.

For a number of instruments, "Cannot Say" responses are given consistently more often with computer administration, unless the testing software is designed to make this response more difficult; this accounts for some (but not all) of the early format differences observed in scale scores (Moreland, 1987). This differential use of the "Cannot Say" response can be controlled for most computer-administered tests¹.

It is difficult to disentangle the effects of repeated testing from format effects when test-retest designs are used in format equivalence research (and most researchers agree that these are the designs of choice), unless fairly complex experimental designs are used (see Schuldberg, in press). One beneficial effect of current research on automated testing is to focus more attention on retest effects in objective personality instruments, issues of person reliability, and more general issues of test occasion equivalence.

Before the era of computerized assessment, format equivalence research was concerned with similarities and differences between, the card, booklet, and tape-recorded, as well as various shortened forms, of the MMPI (see Dahlstrom, Welsh, and Dahlstrom, 1972, pp. 24-28)². To date, relatively little research has been done on individual differences in format effects, their correlates, and

possible underlying psychological processes (Honaker, 1987), despite the fact that a literature exists on individual differences in retest effects. The research reported here uses a variety of techniques to generate measures of both format and retest effects. The experimental design employed provided a partial separation of retest and format effects. However, this design, which used two groups of subjects receiving two forms of the test in counterbalanced order, cannot detect "sensitization" effects for test format, effects related to which form the subject experienced when first exposed to the instrument. In addition, retest effects are partially confounded with format effects in some of the measures of item response instability.

Measures of unsystematic instability in responding.

Previous research on the temporal stability of personality profiles has studied change in either an individual's personality profile or in items across two separate testing occasions (Dahlstrom, Welsh, and Dahlstrom, 1975; Fekken and Holden, 1987; Goldberg, 1978; Goldberg and Jones, 1969; Lewinsohn, 1965; Mauger, 1972; Mills, 1954; Pepper, 1964; Schofield, 1950; Windle, 1954, 1955). Measurement of change in responding at the item level is based on the number of items that are changed between testings³. Indices of the total amount of item shifting are non-directional because they are computed without regard to the direction of change (e.g., "False-True" vs. "True False" vs. "Cannot Say-True", etc.). This paper also refers to such change indices as tapping Unsystematic instability because they do not take into account the properties of the particular items (or scales) that shift for a given subject, or the direction of the

shifts for particular items; the items are not keyed. In addition, if a subject is given two different forms of a test on two different occasions, an unsystematic and non-directional instability index counting total number of changed items combines the shifting attributable to repeated testing and to changes in the test format.

The directional unsystematic shifts in which the subject answered "Cannot Say" in the computerized administration and gave a scoreable "True" or "False" response in the Pencil and paper condition are especially interesting, due to the increased use of the "Cannot Say" response on computerized instruments such as the present one where this response is not limited by the administration software.

Test-retest effects, inconsistent in pattern across studies although sometimes significant in magnitude, have generally been minimized in work with the MMPI. However, an important aspect of research on changes in responding across time involves treating profile or item instability as an individual difference variable and deriving MMPI indices of person reliability. When profile or item instability is treated as a trait, the investigator can then use empirical scale development techniques to derive items that predict test takers' individual levels of the trait. Such a scale may exclude the unstable items themselves (e.g. Pepper, 1964).

The present present paper separates Unsystematic instability into its directional components and computes twelve indices of Unsystematic item instability. These indices count test takers' shifts in item responding between "True", "False", and "Cannot

Say", in each of six possible orders, tallied between formats or across times.

Indices of Systematic profile instability.

The derivation of Systematic instability indices differs from methods for constructing overall or Unsystematic instability scales. Previous research using the same subject pool (Schuldberg, 1987; in press) derived indices of systematic response shifting occurring either between test formats or across repeated testings. A set of forty MMPI items showed significant effects for format of administration but not for repeated testing in item-level crossover analyses of variance (Edwards, 1968; Winer, 1962). These items provide a measure of systematic scale instability because, as a group, subjects' responses to them tend to be different in a particular direction between two testing formats but not across two times. Briefly, these items tend to be scored in opposite directions from and tend to be negatively correlated with the MMPI scales (a notable exception being the K scale), another finding counter to the hypothesis of increased candor in the computerized testing format. A larger number of items (seventy-five) showed significant effects for repeated testing alone. These seventy-five items are used as an index of Systematic time instability⁴.

It was hypothesized that four distinct types of shifting would emerge with different correlates on the MMPI and other measures: 1) Shifting to "Cannot Say" on the computerized administration; 2) Systematic shifting between "True" and "False" responses between test formats; 3) Systematic shifting between

"True" and "False" response across times; and 4) A general index of carelessness reflected in shifts between "True" and "False" both across occasions and between formats.

Methods

Subjects and procedures

The subjects in this research were drawn from a pool of students from an Introductory Psychology class who signed up for a personality assessment study in partial fulfillment of a course experimental requirement. Subjects responding with thirty or more unscorable or "Cannot Say" responses to either form of the test were eliminated. Additional subjects were dropped randomly from the smaller group in order to create two groups of seventy-five subjects each, matched for number of males and females. This resulted in a sample of 150 students, in two groups composed of 42 males and 33 females.

Subjects were tested twice, in counterbalanced order, with two forms of the test given approximately a week apart. The Pencil and paper version of the MMPI presented the group form items in test booklet form. Subjects were tested in a classroom, in groups of up to twenty-five, and were also given the Shipley Institute of Living Scale (Zachary, 1986) after the MMPI. The automated MMPI was administered to groups of students on Apple IIc computers in a microcomputer teaching lab containing 25 machines⁵. After the students' final test administration (either Pencil and paper or Computer), they were given a brief questionnaire about their experience with computers, with objective personality tests, and their perception of the degree

to which the two personality inventories were the same.

Construction of instability indices

Systematic item instability.

An item was selected for one of the two indices of Systematic instability if it showed a relatively unambiguous format or time effect in an item-level crossover ANOVA (Schuldberg, in press). An item with unambiguous format effects showed a significant main effect for format alone (not for time or group). An item showing an unambiguous effect for time was one with a significant main effect for time, but not for format or group. These items were keyed to indicate the direction that -- overall -- the group of 150 test takers shifted on the item either from the first to the second administration or between the Computer and the Paper and pencil administrations. Only shifts involving "True" and "False" responses are included in the two Systematic shifting indices.

For each subject, the number of shifts in the keyed direction for both the format-sensitive and the time-sensitive items was tallied. Each shift in the keyed direction on a selected item was given a weight of +1; each shift opposite to the keyed direction was given a weight of -1; items that did not shift were not counted. This resulted in a score for each subject on two scales: Systematic Format sensitivity and Systematic Time sensitivity.

The internal consistency of the two measures of Systematic instability is low (for Systematic Format instability, Cronbach's $\alpha = 0.25$; for Systematic Time instability, $\alpha = 0.21$), indicating that although these items shift significantly in a particular direction between formats or across times for the

subjects as a group, particular subjects were not consistent in their shifting on these items.

Unsystematic item instability.

Two families of measures of raw, unsystematic item shifting were constructed. These measures are counts of the number of items for which the subject changed his or her response between administrations. Given that a subject answers an item as "True", "False", or "Cannot Say" on two separate occasions, nine sets of responses are possible (see Table 1); six of these represent response shifts. Although order of administration and format are partially confounded in the present design, item shifts can be counted separately for shifts between formats and shifts across occasions.

Insert Table 1 about here

A tally of each of the six types of item shift between Computer and Pencil and paper conditions across all 566 items was made for each subject, as well as a total Unsystematic instability score (the sum of the six instability indicators)⁶. This essentially follows one of Pepper's (1964) strategies for assessing change in responding, although the present study differs from Pepper's in examining specific kinds or directions of item shift (e.g., from "True" to "False")⁷. In the same way, six measures were computed for the shifting of responses to the 566 MMPI items across the first and second occasions of testing. It would be possible to compute an alpha to assess the internal

consistency of each Unsystematic index; however, since each index contains 566 items, this task was beyond the computational capacity available.

Construction of Unsystematic "True-False" indices free of overlapping items for the factor analyses

As described above, the indices of Systematic shifting due to either time or format measure shifting in a keyed direction across times or between formats involving "True" and "False" responses. The items on these indices are all included on the Unsystematic "True-False" and "False-True" shifting indices tallied between formats or across times for all 566 items. For the factor analytic studies, special indices of Unsystematic item shifting between "True" and "False" (and "False" and "True") responses between formats and across times were constructed that excluded the items on the Systematic format or Systematic time indices. While these non-overlapping Unsystematic indices are highly correlated with the basic indices computed over all 566 items, they are used in the factor analyses in order to reduce spurious correlation among the variables introduced by item overlap.

Other measures

Relationships between the indices of Systematic and Unsystematic item instability and profile validity (assessed by the four MMPI validity scales), as well as subjects' personality characteristics (measured by the MMPI clinical scales), scores on the Shipley Institute of Living Scale (Zachary, 1986), age, and computer experience are examined.

Computer and related experience was measured using responses to three seven-point Likert-type items: "How much experience with computers have you had before this experiment?" (rated from "no experience" to "a great deal of experience"), "How often do you type or use a typewriter keyboard (such as on a computer terminal)?" (rated from "never" to "every day"), and "How often do you play with video games or computer games?" (also rated from "never" to "every day"). The average of the responses to these three items served as an index of computer and related experience. Cronbach's alpha for this three-item scale is 0.61.

Analyses

Correlates of each of the two Systematic instability indices, the twelve Format and Time Unsystematic instability indices, and the total Unsystematic instability index are examined and discussed. Two Principal-components factor analyses with Varimax rotation are used to examine the factors underlying the item instability indices. Since the Unsystematic indices of item shifting confound the effects of time and administration format, separate analyses are conducted for the Format and Time indices, in each case including the Systematic Instability index and the six format or time Unsystematic instability indices.

Results

Subjects shifted most often between "True" and "False" responses (or vice-versa), changing an average of 43.8 items in either direction. On average, the test takers made shifts of any kind in their responses to 94.1 items, higher than the value of 67.0 reported by Fekken and Holden (1987) for shifting across

repeated testings alone. The Total Unsystematic shifting index ranged from 32 to 249 items.

Table 2 presents the correlations of the Format instability indicators, both Systematic and Unsystematic, with the MMPI scales and total Shipley score. The Systematic Format instability index is significantly negatively correlated only with MMPI Scale 3 (Hysteria) and the Shipley score, as well as showing weak positive correlations with several MMPI scales reflecting deviant traits. The Unsystematic index of shifting

Insert Table 2 about here

from "True" in the Computerized administration to "False" in the Pencil and paper condition is positively correlated with F and seven of the clinical scales, as well as being negatively correlated with L, K, Scale 3 again, and the Shipley score. Correlations for the Unsystematic shifting indices involving "Cannot Say" are mainly related to "Cannot Say scores" themselves, with small negative correlations occurring for the K scale (for "Cannot Say-True" shifts) and Shipley score. The total Unsystematic shifting index (consisting of the sum of all item shifts between administrations) is positively correlated with "Cannot Say" and F, negatively correlated with K, a finding consistent with other research (Fekken and Holden, 1987; Windle, 1954), highly negatively correlated with the Shipley score, and positively correlated with four clinical scales.

A similar pattern of results emerges for the Unsystematic indices of item shifting across times (See Table 3). The

Systematic Time Instability index, computed on the basis of keyed

 Insert Table 3 about here

item shifting on the seventy-five items showing significant effects for Time in earlier crossover ANOVA's, is negatively correlated with L ($r = -0.30$, $p < 0.001$). Unsystematic "True-False" shifting from first to second administration is positively correlated with F and four clinical scales, and negatively correlated with the K scale and total Shipley score. Shifting involving "Cannot Say" is, of course, related to raw "Cannot Say" scores, as well as showing small correlations in different directions on two clinical scales. None of the instability indices is correlated with age or the measure of computer and related experience.

Factor analyses were conducted to examine the relationships among various instability indices. The first factor analysis includes the Format Instability indicators and the second the Time Instability Indices; these analyses were conducted separately, due to the fact that the Unsystematic indices of Time and Format shifting are confounded. In each factor analysis, the special indices of Unsystematic "True-False" and "False-True" shifting described above are used; the items keyed "True-False" or "False-True" on the Systematic shifting index entering in the same analysis are excluded. This eliminates any relationships among the variables in a given analysis solely due to common items. A three-factor solution was derived in each analysis.

The first factor analysis included the Systematic Format

shifting index and the six indices of Unsystematic shifting across Formats. Three factors accounted for 78% of the variance in these measures. The first factor includes the measures that refer to use of the "Cannot Say" response in the Pencil and paper Format. The second factor contains the two indices involving the "Cannot Say" response in the computerized test format. The third factor contains the two Unsystematic indices of shifting between "True" and "False" responses between formats (both "True-False and "False-True"), as well as the Systematic Format Index, which also refers to shifting between "True" and "False" and vice versa.

Similar results are obtained in the second factor analysis, which includes the Systematic and Unsystematic Time indices. Three factors accounted for 74% of the variance in this analysis. Again, two "Cannot Say" factors emerge (one for Time 1 and one for Time 2), along with a factor on which the Unsystematic "True-False" shifting indices load. Unlike the Systematic Format Instability index, the Systematic Time Instability index shares

Insert Tables 4 and 5 about here

virtually no common variance with the three factors.

Discussion

With regard to the specific hypotheses of this study, it does appear that use of the "Cannot say" response is different under the two test formats. However, there do not appear to be several distinct varieties of "True-False" shifting. This type of response alternation emerges as a unitary phenomenon,

regardless of the direction of the shift or whether the particular items involved are "format sensitive" or not. This "True-False" shifting appears to be related to a more deviant (or careless) response set, and co-occurs with more elevated overall profiles. The Systematic Time shifting scale, on the other hand, is unrelated to the other indices in this study.

The measures of Systematic instability due to Format or Time have low alpha 's, indicating that test takers do not shift their responses to a particular consistent set of MMPI items, either between administration formats or across occasions. Neither index taps a consistent pattern of individual differences in response shifting. The systematic tendency to shift items between "True" and "False" across time was unrelated to the other instability measures, and may possibly be related to a general tendency to respond to the test with candor; the Systematic Time shifting indicator is negatively correlated with the MMPI "Lie" scale. Although research has suggested that response shifting on retest may occur in the direction of increased Social Desirability (Windle, 1954), the present finding indicates that the particular subjects who shift in this way may be more candid ones.

This study finds at least three different types of instability in item responding. The first is a general tendency for some test takers to shift between "True" and "False" responses when taking the MMPI twice. It is unclear whether this shifting is primarily attributable to time or format, although it appears doubtful on the basis of previous research that some test takers shift between "True" and "False" responses on the basis of the format of the test alone; when retest involves a different test

format, however, the average number of items changed is greater than reported elsewhere in the literature. The general tendency to shift between "True" and "False" responses to the test may reflect invalid or "unreliable" responding, as it is associated with elevated scores on F and more generally elevated profiles. The negative correlations between the "True-False" shifting indices and the Shipley score may indicate that this represents an "intellectually easy" approach to the test or reflects a cognitive deficit. However, these negative correlations may also indicate that test takers who made a large number of "True-False" shifts (or vice versa) were poorly motivated in the testing situation and responded to the Shipley and at least one of the MMPI's in a haphazard fashion.

The use of the "Cannot Say" response contains two underlying factors, related either to Format or Time of administration⁸. The two factor analyses taken together tend to muddy the effects of Format and Time on "Cannot Say" scores. However, as a group, the test takers' use of the "Cannot Say" response shows a highly significant effect for Format and not for Time (Schuldburg, in press); this indicates that the two factors underlying the "Cannot Say" response are "Computer format" and "Pencil and paper format", as emerged in the first factor analysis. The two "Cannot Say" factors from the second factor analysis appear to be an artifact of the confounding of format and time effects. The main finding of the second factor analysis is that the Systematic Time shifting index, computed from each subject's keyed shifting on the seventy-five items showing significant time effects, is unrelated to the other indices of response variability in this

study. Although this Systematic Time index has low internal consistency, it taps a distinct domain of response tendencies not measured by the other indices in this study. The present research suggests the importance of continued awareness and renewed examination of the effects of repeated testing on subjects' MMPI responses.

In conclusion, the shifting of item responses between two different test formats appears to be an inconsistent although measurable phenomenon, mainly reducible to "True-False" shifting and to differential use of "Cannot Say" in some computerized tests. The "Cannot Say" response appears to represent a separate phenomenon in computerized and traditional MMPI administration, and is sensitive to the design of the administration software. Systematic Time shifting is a more substantial and robust effect than Format shifting; however, its relationship to other person variables remains something of a mystery.

References

- Ben-Porath, Y. S., & Butcher, J. N. (1986). Computers in personality assessment: A brief past, an ebullient present, and an expanding future. Computers in Human Behavior, 2, 167-182.
- Biskin, B. H., & Kolotkin, R. L. (1977). Effects of computerized administration on scores on the Minnesota Multiphasic Personality Inventory. Applied Psychological Measurement, 1, 543-549.
- Bresolin, M.J., Jr. (1984). A comparative study of the computer administration of the Minnesota Multiphasic Personality Inventory in an inpatient psychiatric setting. Unpublished doctoral dissertation, Loyola University, Chicago.
- Butcher, J.N. (1987). The use of computers in psychological assessment: An overview of practices and issues. In J.N. Butcher (Ed.), Computerized psychological assessment: A practitioner's guide (pp. 3-14). N.Y.: Basic Books.
- Dahlstrom, W.G., Welsh, G.S., & Dahlstrom L.E. (1972). An MMPI handbook. Volume 1: Clinical interpretation. Minneapolis: University of Minnesota Press.
- Dahlstrom, W.G., Welsh, G.S., & Dahlstrom L.E. (1975). An MMPI handbook. Volume 2: Research Applications. Minneapolis: University of Minnesota Press.
- Edwards, A. L. (1968). Experimental Design in Psychological Research (Third Edition). N.Y.: Holt, Rinehart, & Winston.
- Edwards, A., & Walsh, J.A. (1964). A factor analysis of 7 scores. Journal of Abnormal and Social Psychology, 69, 559-563.

- Evan, W. M., & Miller, J. R., III. (1969). Differential effects on response bias of computer vs. conventional administration of a social science questionnaire: An exploratory methodological experiment. Behavioral Science, 14, 216-227.
- Fekken, G. C., & Holden, R. R. (1987). Assessing the person reliability of an individual MMPI protocol. Journal of Personality Assessment, 51, 123-132.
- Goldberg, L. R. (1978). The reliability of reliability: The generality and correlates of intra-individual consistency in responses to structured personality inventories. Applied Psychological Measurement, 2, 269-291.
- Goldberg, L. R., & Jones, R.R. (1969). The reliability of reliability: The generality and correlates of intra-individual consistency in responses to structured personality inventories. Oregon Research Institute Research Monographs, 9 (2).
- Goldberg, L. R., & Rust, R. M. (1964). Intra-individual variability of the MMPI-CPI common item pool. British Journal of Social and Clinical Psychology, 3, 145-147.
- Harris, W. G., & Allred, L. J. (1987). Computer administration of the Adjective Check List: A study of mode effect. Paper presented at the 95th. annual meeting of the American Psychological Association, New York, August, 1987.
- Hart, R. R., & Goldstein, M. A. (1985). Computer-assisted psychological assessment. Computers in Human Services, 1, 69-75.
- Hathaway, S.R. & McKinley, J.C. (1967). Minnesota Multiphasic Personality Inventory: Manual (Revised edition). New York: The Psychological Corporation.

- Holden, R. R., Helmes, E., Fekken, G. C., & Jackson, D. N. (1985). The multidimensionality of person reliability: Implications for interpreting individual test item responses. Educational and Psychological measurement, 45, 119-130.
- Honaker, L.M. (1987). The equivalency of computerized and conventional tests: Current status. Symposium presented at the 95th. annual meeting of the American Psychological Association, New York, August, 1987.
- Koson, D., Kitchen, C., Kochen, M., & Stodolosky, D. (1970). Psychological testing by computer: Effect on response bias. Educational and Psychological Measurement, 30, 803-810.
- Lambert, M.E., Andrews, R.H., Rylee, K., & Skinner, J.R. (1987). Equivalence of computerized and traditional MMPI administration with substance abusers. Computers in Human Behavior, 3, 139-143.
- Layton, W. L. (1954). The variability of individuals' scores upon successive testings on the Minnesota Multiphasic Personality Inventory. Educational and Psychological Measurement, 14, 634-640.
- Lewinsohn, P. M. (1965). Dimensions of MMPI change. Journal of Clinical Psychology, 21, 37-43.
- Lukin, M. E., Dowd, E. T., Plake, B. S., & Kraft, R. B. (1985). Comparing computerized versus traditional psychological assessment. Computers in Human Behavior, 1, 49-58.
- Lushene, R. E., O'Neil, H. F., & Dunn, T. (1974). Equivalent validity of a completely computerized MMPI. Journal of Personality Assessment, 38, 353-361.

- Mauger, P.A. (1972). The test-retest reliability of persons: An empirical investigation using the MMPI and Personality Research Form. Doctoral dissertation, University of Minnesota. (Dissertation Abstracts International, 33, 2816B).
- Mills, W.W. (1954). MMPI profile pattern and scale stability throughout four years of college attendance. Doctoral dissertation, University of Minnesota (Dissertation Abstracts International, 14, 1259).
- Moreland, K.L. (1987). Computerized psychological assessment: What's available. In J.N. Butcher (Ed.), Computerized psychological assessment: A practitioner's guide, (pp. 3-14). N.Y.: Basic Books.
- Pepper, L.J. (1964). The MMPI: Initial test predictors of retest changes. (Dissertation Abstracts International, [1965], 26, 1780). (Doctoral Dissertation, University of North Carolina.)
- Rezmovic, V. (1977). The effects of computerized experimentation on response variance. Behavior Research Methods and Instrumentation, 9, 144-147.
- Russell, G. K. G., Peace, K. A., & Mellsop, G. W. (1986). The reliability of a micro-computer administration of the MMPI. Journal of Clinical Psychology, 42, 120-122.
- Schofield, W. (1950). Changes in responses to the MMPI following certain therapies. Psychological Monographs, 64, 5 (Whole no. 311).

- Schuldberg, D. (1987). The human-machine "relationship" in computer-assisted MMPI administration: Sensitivity of MMPI items to the automated testing format. Paper presented to the 22nd. Annual Symposium on Recent Developments in the Use of the MMPI, Seattle, Wa., May 2-3, 1987.
- Schuldberg, D. (1988). The MMPI is less sensitive to the automated testing format than it is to repeated testing: Item and scale effects. Computers in Human Behavior, 4, 285-298.
- Space, L. G. (1981). The computer as psychometrician. Behavior Research Methods and Instrumentation, 13, 595-606.
- Winer, B.J. (1962). Statistical principles in experimental design. N.Y.: McGraw-Hill.
- Wilson, F.R., Genco, K., & Yager, G.G. (1985). Assessing the equivalence of paper and pencil vs. computerized tests: Demonstration of a promising methodology. Computers in Human Behavior, 1, 265-275.
- White, D. M., Clements, C. B., & Fowler, R. D. (1985). A comparison of computer administration with standard administration of the MMPI. Computers in Human Behavior, 1, 153-162.
- Windle, C. (1954). Test-retest effect on personality questionnaires. Educational and Psychological Measurement, 14, 617-633.
- Windle, C. (1955). Further studies of test-retest effect on personality questionnaires. Educational and Psychological Measurement, 15, 246-253.
- Zachary, R.A. (1986). Shipley Institute of Living Scale: Revised manual. Los Angeles: Western Psychological Services.

Footnotes

¹However, responding to certain types of tests may be fundamentally changed by the automated format. For example, the Adjective Check List may become more like a forced choice instrument involving paired comparisons for each adjective in the automated format, resulting in very different patterns of responding (see Harris and Allred, 1987).

²This research turned up significant but inconsistent patterns of format differences.

³Items that recur within the test can also be examined for consistency within a single occasion of testing.

⁴Items showing effects for group or effects for a combination of time, group, and format were discarded as exhibiting ambiguous and uninterpretable effects. The composition of the Systematic Format sensitivity index is as follows: Items keyed True ($\underline{n} = 55$): 3, 7, 8, 9, 36, 46, 68, 73, 79, 88, 96, 98, 99, 118, 119, 128, 130, 131, 133, 146, 155, 160, 164, 175, 188, 221, 222, 228, 229, 230, 254, 261, 264, 274, 277, 302, 304, 306, 307, 318, 346, 369, 372, 376, 377, 379, 381, 428, 434, 435, 495, 522, 523, 537, 542. Items keyed false ($\underline{n} = 20$): 16, 22, 62, 71, 84, 90, 93, 134, 215, 217, 297, 313, 332, 334, 390, 397, 436, 497, 543, 560. (Numbers refer to the booklet form of the MMPI.) The keying of these items is based on the direction of group changes from the first to the second administration. An item is keyed "True" when it was answered significantly more often as "False" on the first administration and "True" on the second. An item keyed "False" was answered significantly more often as "True" on the first administration and "False" on the second.

⁵The items were presented in Form R order in the computer administration condition and in booklet form order in the Pencil and paper condition, representing a confound of test format and order of the items in the present study.

⁶Only one total Unsystematic item instability score was computed because the overall, total Unsystematic Instability measure collapses Format and Time effects; the value of the total Unsystematic instability measure is identical whether it is computed using the individual Format or Time indices.

⁷In contrast to Pepper's (1964) work, the present research does not construct a separate scale for predicting subjects' scores on the shifting indices.

⁸These findings may be compared to Edwards and Walsh's (1964) analysis of "Cannot Say" scores. Examining responses from Pencil and paper tests, these authors found three interpretable factors underlying "Cannot Say" responses, in particular that "Cannot Say" responses are different for True-False and forced-choice items. A similar distinction may be useful in classifying different type of automated assessment formats (see footnote 1).

Author Notes

I am grateful to James A. Walsh for his continuing help with the methodology used in this paper, and to the helpful comments of Bruce H. Biskin. I wish to acknowledge the assistance of the University of Montana Psychology Department, and grants from Montanans on a New Track for Science (MONTTS) and the University of Montana Office of Research Administration

Table 1.

Systematic and Unsystematic Instability Indices.

Systematic Instability Indices.

Format: 40 items showing significant response differences in a particular direction attributable to format.

Time: 75 items showing significant response differences in a particular direction attributable to time.

Unsystematic Item shifts tallied between Formats and across Times

Instability indices:

T-F

T-?

F-T (These shifts are assessed both
F-? between formats and across times.)

?-T

?-F

Stable response combinations:

T-T

F-F

?-?

Table 2.

Correlation of Format instability scales with MMPI scales and Shipley.

Measure	Instability Scale							
	Systematic Format Instability	Unsystematic Format Instability						Total Unsystematic Instability ^b
		T-F	T-?	F-T	F-?	?-T	?-F	
MMPI scales								
?	-.13	-.06	.11	.08	.16	.89***	.92***	.22**
L	.09	-.20*	-.05	.04	.01	-.04	.08	-.09
F	.15	.42***	.04	.16	.05	.10	.01	.35***
K	-.10	-.50***	-.01	-.14	.01	-.17*	-.08	-.40***
1	.03	.26***	.12	.24***	.11	.07	.03	.31***
2	-.03	.06	-.01	.03	.10	-.08	-.09	.04
3	-.18*	-.22**	.11	.08	.06	-.00	-.05	-.09
4	-.00	.19**	.14	.10	.12	-.02	.00	.18*
5	-.03	.08	.12	-.04	.04	-.11	-.16	.00
6	.01	.17*	.01	.01	-.04	-.04	-.07	.09
7	.06	.47***	.07	.13	.02	.08	-.01	.36***
8	.16	.49***	.07	.17*	.02	.11	.03	.41***
9	.15	.40***	.02	.08	-.07	.00	.02	.29***
0	.12	.19*	-.03	.09	.04	-.04	-.10	.15
Shipley ^a	-.17*	-.41***	-.15	-.43***	-.02	-.14	-.17*	-.52***

* $p \leq .05$ ** $p \leq .01$ *** $p \leq .005$ Two-tailed tests of significance.

Note: MMPI scale scores are taken from the computerized administration of the MMPI. Results were similar for scores obtained from the Pencil and paper version of the MMPI.

^aTotal raw score on the Shipley Institute of Living Scale.

^bDue to the fact that format and time of administration are not independent factors in this study, the total Unsystematic instability index, computed as the sum of all six Unsystematic instability indices, is the same regardless of whether the individual format or time measures are used.

Table 3.
Correlation of Time instability scales with MMPI and Shipley.

measure	Instability Scale						
	Systematic Time Instability	Unsystematic Time Instability					
		T-F	T-?	F-T	F-?	?-T	?-F
MMPI scales							
?	.06	-.06	.58***	.06	.59***	.56***	.61***
L	-.30***	-.09	.05	-.08	.13	-.12	-.04
F	.07	.24***	.00	.35***	-.03	.14	.07
K	-.02	-.27***	-.08	-.39***	.04	-.13	-.15
1	.00	.23***	.02	.28***	-.08	.13	.19*
2	.04	.06	-.14	.04	-.13	.03	.09
3	.05	-.08	-.05	-.09	-.10	.10	.08
4	-.03	.16	-.09	.15	-.05	.13	.12
5	-.12	.01	-.04	.04	-.08	-.02	-.09
6	.01	.06	-.17*	.13	-.10	.11	.00
7	.08	.26***	.00	.36***	-.11	.14	.12
8	.09	.26***	.04	.41***	-.08	.14	.14
9	.05	.24***	-.01	.26***	-.00	.02	-.01
0	.02	.11	-.10	.17*	-.17*	.03	.08
Shipley ^a	-.08	-.43***	-.13	-.42***	-.05	-.13	-.17*

* $p \leq .05$ Two-tailed tests of significance.

** $p \leq .01$

*** $p \leq .005$

Note: MMPI scale scores are based on the computerized administration of the MMPI. Results were similar for scores obtained from the Pencil and paper version of the MMPI.

^aTotal raw score on the Shipley Institute of Living Scale.

Table 4.

Factor Analysis of the item instability measures: Format

Response Measure	Factor 1 "Cannot Say: P and P"	Factor 2 "Cannot Say: Computer"	Factor 3 "TF/FT Shiftiness"	Communality
Systematic Shift				
Format	-0.08	-0.13	<u>0.69</u>	0.50
Unsystematic shifts (Computer to Pencil and Paper):				
T-F ^a	-0.10	0.02	<u>0.79</u>	0.63
T-?	0.99	-0.01	<u>-0.05</u>	0.98
F-T ^a	-0.06	0.04	<u>0.79</u>	0.63
F-?	0.99	-0.00	<u>-0.07</u>	0.98
?-T	-0.00	<u>0.93</u>	0.05	0.88
?-F	-0.01	<u>0.93</u>	-0.07	0.87
Percent variance accounted for	0.28	0.25	0.25	Total 0.78

^aThese Unsystematic T-F and F-T indicators exclude the MMPI items on the Systematic Format shifting index.

Note:

Loadings for the variables used in naming the factor are underlined.

Table 5.
Factor Analysis of the item instability measures: Time

Response Measure	Factor 1 "Cannot Say: Time 1"	Factor 2 "Cannot Say: Time 2"	Factor 3 "TF/FT Shiftiness"	Communality
Systematic Shift				
Time	-0.26	-0.30	-0.04	0.16
Unsystematic Shifts (First to second administration):				
T-F ^a	-0.06	-0.10	<u>0.84</u>	0.71
T-?	-0.04	<u>0.93</u>	-0.05	0.86
F-T ^a	0.02	0.00	<u>0.86</u>	0.74
F-?	-0.04	<u>0.93</u>	-0.03	0.87
?-T	<u>0.96</u>	-0.03	0.03	0.93
?-F	<u>0.96</u>	-0.03	-0.05	0.92
Percent variance accounted for	0.27	0.26	0.21	total 0.74

^aThe Unsystematic T-F and F-T indicators exclude the MMPI items on the Systematic Time shifting index.

Note:

Loadings for the variables used in naming the factor are underlined.